Journal of Vegetation Science 12: 15-26, 2001
© IAVS; Opulus Press Uppsala. Printed in Sweden

Modeling spatially explicit forest structural attributes using Generalized Additive Models

Frescino, Tracey S.[1,2]; Edwards, Thomas C., Jr.[3] & Moisen, Gretchen G.[1]

[1]*USDA Forest Service, Rocky Mountain Research Station, 507 25th Street, Ogden, UT 84401, USA;*
[2]*Graduate Degree Program in Fisheries and Wildlife, Utah State University, Logan, UT 84322-5210, USA;*
[3]*USGS Biological Resources Division, Utah Cooperative Fish and Wildlife Research Unit, Utah State University, Logan, UT 84322-5210, USA;* **Corresponding author; E-mail tfrescino@fs.fed.us*

Abstract. We modelled forest composition and structural diversity in the Uinta Mountains, Utah, as functions of satellite spectral data and spatially-explicit environmental variables through generalized additive models. Measures of vegetation composition and structural diversity were available from existing forest inventory data. Satellite data included raw spectral data from the Landsat Thematic Mapper (TM), a GAP Analysis classified TM, and a vegetation index based on raw spectral data from an advanced very high resolution radiometer (AVHRR).

Environmental predictor variables included maps of temperature, precipitation, elevation, aspect, slope, and geology. Spatially-explicit predictions were generated for the presence of forest and lodgepole cover types, basal area of forest trees, percent cover of shrubs, and density of snags. The maps were validated using an independent set of field data collected from the Evanston ranger district within the Uinta Mountains. Within the Evanston ranger district, model predictions were 88% and 80% accurate for forest presence and lodgepole pine (*Pinus contorta*), respectively. An average 62% of the predictions of basal area, shrub cover, and snag density fell within a 15% deviation from the field validation values. The addition of TM spectral data and the GAP Analysis TM-classified data contributed significantly to the models' predictions, while AVHRR had less significance.

Keywords: Accuracy assessment; AVHRR; Forest attribute model; Generalized additive model; Geographical Information Systems; Landsat Thematic Mapper; Vegetation modelling.

Abbreviations: AIC = Akaike's Information Criterion; AVHRR = Advanced Very High Resolution Radiometer; DBH = Diameter Breast Height; DMA = Defense Mapping Agency; GAM = Generalized Additive Model; GLM = Generalized Linear Model; FIA = Forest Inventory and Analysis; GIS = Geographical Information Systems; GPS = Global Positioning System; PCC = Percent correctly classified; RMS = Root Mean Square error; TM = Thematic Mapper; UTM = Universal Transverse Mercator.

Introduction

Recent advances in statistical modelling techniques and geographical tools, such as remote sensing and geographical information systems (GIS), have increased the opportunities for the delineation and analysis of vegetation distribution patterns. Numerous studies have demonstrated the use of statistical models to understand and display how plant species are distributed throughout the environment (e.g. Austin et al. 1990; Davis & Goetz 1990; Austin et al. 1994), yet the unpredictability of natural ecosystems, along with the dramatic influence of human disturbance, has made it very difficult to draw conclusions about vegetation distribution patterns and relationships to environmental conditions. For example, research has demonstrated that the past assumption that vegetation responds in a bell-shaped (Gaussian) pattern along environmental gradients is not true for most species (Mueller-Dombois & Ellenberg 1974; Austin & Cunningham 1981; Austin 1987). Many statistical models being applied to vegetation hold this assumption and therefore tend to misrepresent true distributional patterns (e.g., ordination methods; Austin & Noy-Meir 1971; Austin 1985). Other statistical models, such as generalized additive models (GAMs), are more flexible and better suited to handle nonlinear relationships of vegetation to environmental gradients (Hastie & Tibshirani 1990; Yee & Mitchell 1991; Austin & Meyers 1996).

GIS and remote sensing technology have made it possible to identify, analyze, and classify extensive tracts of vegetation using satellite spectral data and digital environmental data. Studies have shown the complementary effects of integrating environmental data with satellite spectral data for vegetation classification (Loveland et al. 1991; Homer et al. 1997), stratification (Franklin et al. 1986) and predictive modelling (Frank 1988; Davis & Goetz 1990; Moisen & Edwards 1999). GIS tools allow such integration, storage, and spatial analysis of multiple layers of data and provide methods for generating georeferenced maps. When analyzing large areas, questions arise whether to use a readily available satellite data source, such as 1.1km resolution, National Oceanic and Atmospheric Administration's (NOAA) advanced very high resolution radiometer (AVHRR) or a higher resolution data source, such as 30-m, multi-spectral, Landsat Thematic Mapper (TM) imagery which is more expensive and requires extensive storage space.

Although the development of large-scale analytical tools has increased efficiency, most research has focused on dominant vegetation features that are distinguishable from satellites or that represent climax or seral types most influenced by environmental parameters. But how do we analyze the understory and composition of forested habitats that are not directly visible from satellites? Studies have looked at the ability of satellites to capture reflectance values of understory components (Stenback & Congalton 1990), basal area and leaf biomass (Franklin 1986), and stand density and height (Horler & Ahern 1986), but in general, further research was suggested.

This study outlines an approach for delineating forest composition using GAMs, remote sensing data, and GIS tools. Our overall objective was to determine the ability of these techniques, when integrated, to model and map attributes of forest structure in the Uinta Mountains of Utah, and at the same time develop a systematic approach for application of these techniques to other forested landscapes. Specifically, our objectives were to:

(1) develop spatially explicit predictive models of forest attributes using GAMs, integrating field-collected forest resource inventory data with satellite and digital environmental data;

(2) determine the effects of three different forms of imagery (Landsat TM, AVHRR, and a classified TM-based vegetation cover map) on model predictive capabilities; and

(3) test how well the models predict at a local level using an independent set of field data.

Methods

Study area

Data for model-building came from a region of seven National Forest Ranger Districts encompassing the east-west mountain range of the Northern Utah Mountain Ecoregion (hereafter the Uinta mountains). The seven ranger districts together cover approximately 1 000 000 ha of forest. The Uintas are characterized by an east-west orientation, and have an approximate length of 241 km and a width of 48 to 64 km. Elevation ranges from ~1700 m to a high of ~4000m. The area contains conspicuously deep, v-shaped canyons on the south side of the range and less pronounced canyons on the north side of the range. The geology consists mainly of a sedimentary layer of sandstone and limestone in the forested areas, glacial deposits in the valleys and drainages, and Precambrian quartzite in the high elevation, exposed regions. The climate consists of long winters and high summer precipitation which is mainly a function of elevation, latitude, and storm patterns from the west and the Gulf of Mexico, with local effects from slope exposure and/or aspect (Mauk & Henderson 1984).

The distribution of vegetation in the Uinta Mountains is highly influenced by topographic position and geographic location. *Pinus contorta* (Lodgepole pine) is the dominant vegetation type, ranging from 1700 to 3000 m elevation. At elevations between 2400 m and 3000 m, lodgepole is mixed with *Populus tremuloides* (aspen), with a few homogenous aspen stands at lower elevations. As elevation increases, lodgepole forests are gradually replaced by *Picea engelmannii-Abies lasiocarpa* (spruce-fir) forest types and are frequently interspersed with large patches of wet and dry meadows. Other forest types include *Pinus edulis-Juniperus osteosperma* (pinyon-juniper) at lower elevations on the northeastern slope, *Pseudotsuga menziesii* (Douglas-fir) on steep, protected slopes, and *Pinus ponderosa* (ponderosa pine) forests on exposed slopes on the south side of the range (Cronquist et al. 1972). Human impacts on natural successional processes within the Uintas include timber management and wood collection, fire suppression, intensive grazing, recreation, and intensive harvesting of lodgepole pine forests for railroad tie (= railway sleeper) production in the early 1900's.

Data

Response variables

Forest attribute data were extracted from the U.S. Forest Service Rocky Mountain Research Station, Interior West Resource Inventory, Monitoring, and Evaluation Program (IWRIME) database (Anon. 1994). Five forest attributes were chosen as response variables for this study: two binomial (forest presence and *Pinus contorta* presence) and three continuous variables (live basal area, percent shrub cover, and snag density) (Table 1). Forest was defined as land, 0.4 ha or more in size, having at least 10% tree cover. A location was classified as lodgepole forest type when the majority of tree cover in a forested site was lodgepole. Live basal area was calculated from measured diameter at breast height

(DBH) of timber trees 2.5 cm or greater DBH, and a sum of diameter at root collar for woodland trees > 7.6 cm. Percent shrub cover was derived from total shrub cover of three different height classes, calculated by summing the midpoints of each specified cover class (< 5%, 5-25%, 25-50%, 50-75%, or 75-100%) measured in the field. Snag density was a measure of salvable and nonsalvable timber snags greater than 10.2 cm DBH, per 0.4 ha plot. Snags were counted within a 25.3 m radius and multiplied by 2 for a 0.4 ha estimate. For further information on FIA sampling and measurement procedures, accuracy standards, and other sampled parameters, refer to USDA (Anon. 1994).

Explanatory variables

The selection of explanatory variables for modelling was based on a priori ecological assumptions and published literature on vegetation responses to environmental gradients, and the availability of appropriate digital coverages within the study area. Each initial model included total annual precipitation, three topographic variables (elevation, aspect, and slope), geology, three geographical location variables (UTM easting and northing coordinates and a discrete variable of ranger district), and one of three types of satellite spectral data (AVHRR, Landsat TM, or a classified Landsat TM-based vegetation cover map) (Table 2).

Precipitation data came from a downscaling of coarse-scale Prism (Daly et al. 1994) climate maps (N. Zimmermann, unpubl. data). Elevation was extracted from the Defense Mapping Agency (DMA), 90-m resolution, digital elevation models. Aspect and slope data were derived from the DMA using functions in the GRID module of ArcInfo GIS (ESRI Inc., Redlands, California). From aspect, azimuth in degrees was transformed into three different variables. The first variable (Asp1) was derived from a look-up table of slope and aspect providing estimates of relative total annual solar radiation normalized at 41 degrees latitude (Swift 1976). The second variable (Asp2) was a discrete variable separated into categories of degrees. The categories range from 1 to 9, with category 1 as north-facing aspect, moving clockwise to category 8 at northwest aspects. Category 9 included slopes less than five percent. The third aspect variable (Asp3) was a symmetric radiation wetness index transformed from aspect degrees (Roberts & Cooper 1989).

Geology data were obtained from a digitized coverage of a 1:500000 stable base mylar of the geology of Utah (Hintze 1980). Three groups of discrete variables were derived from the geology coverage by combining features into classes based on nutrient quality (1-sandstone and limestone, 2-sedimentary, 3-alluvial), time era (1-Precambrian, 2-Mississippian to Euocene, 3-Alluvium), and rock type (1-sedimentary, 2-alluvial) (see Frescino 1998: Appendix A1).

Geographic location was represented by the Universal Tranverse Mercator (UTM) easting and northing values. The last explanatory variable was a discrete variable with seven components representing the seven National Forest Ranger Districts (1-Evanston, 2-Mountain View, 3-Flaming Gorge, 4-Vernal, 5-Roosevelt, 6-Kamas, 7-Duchesne). Although not ecologically defined, the districts have characteristic boundaries which are associated with geographical features.

Three types of satellite data were compared in this study: TM-based classified imagery; AVHRR; and unclassified Landsat TM. The TM-based, classified map of 36 classes was developed from a georeferenced mosaic of TM scenes (see Homer et al. 1997 for details). For this study, these 36 classes were reclassified to match IWRIME forest type classes, resulting in a total of 8 categories (Frescino 1998: Appendix A3). A binary variable of forest and non-forest types was also classified for use in the model predicting forest presence/absence. The AVHRR data source used was the normalized difference vegetation index (hereafter AVHRR) (Loveland et al. 1991). The third type of satellite data was unclassified TM. Only bands 3 (Red), 4 (Near-infrared), and 5 (Mid-infrared) were used in the TM-based models. Visible bands, 1 and 2, and mid-infrared band 7 were highly correlated with bands 3, 4, and 5, and were removed from the analysis.

Each digital coverage was rescaled within the GIS to a cell size of 0.4 ha using the cubic convolution algo-

Table 1. Summary of response variables for modeling forest attributes in the Uinta Mountains, Utah. Data collected from 0.4 ha-size plots following procedures and definitions in Anon. 1994.

Forest attribute	Type	Description	Distribution
Forest presence	Binomial	> 10% stocking (> 61 m wide)	Pr = 0.77
Presence of *Pinus contorta*	Binomial	Majority of forest cover	Pr = 0.31
Basal area (m^2/ha)	Continuous	Area of trees at 1.37 m basal (trees > 2.5 cm DBH)	Range: 0 to 70 Median: 16
Shrubs (%)	Continuous	Sum of total cover from upper, mid, and lower layers	Range: 0 to 92 Median: 15
Snag density	Continuous	Total salvable and non-salvable (snags > 10.2 cm DBH)	Range: 0 to 248 Median: 5

Table 2. Summary of explanatory variables used to model forest attributes in the Uinta Mountains, Utah, USA.

Variable	Abrev.	Type	Resolution	Source
Elevation(m)	Elev	Continuous	90 m	DMA
Asp (°)	-	-	-	Derived from DMA
	Asp1	Continuous	90 m	Relative annual solar radiation (Swift 1976)
	Asp2	Discrete	90 m	9 categories (see text for descriptions)
	Asp3	Continuous	90 m	Radiation/wetness index (Roberts & Cooper 1989)
Slope(%)	Slp	Continuous	90 m	Derived from DMA
Precipitation (mm)	Precip	Continuous	90 m	Downscaled from PRISM-yearly precipitation climate maps (N. Zimmerman, unpubl. data)
Geology	-	-	-	Hintze (1980)
	Geol(T)	Discrete	1:500000	Timeframe (1-Precambrian, 2-Mississippian to Euocene, 3-Alluvium)
	Geol(N)	Discrete	1:500 000	Nutrients (1-sandstone and limestone, 2-sedimentary, 3-alluvial)
	Geol(R)	Discrete	1:500,000	Rock Type (1-sedimentary, 2-alluvial)
Easting	East	Continuous	-	UTM Easting coordinates
Northing	North	Continuous	-	UTM Northing coordinates
District	District	Discrete	-	National Forest Ranger Districts (1-Evanston, 2-Mountain View, 3-Flaming Gorge, 4-Vernal, 5-Roosevelt, 6-Kamas,7-Duchesne)
TM-classified	GAPveg	Discrete	90 m	GAP Analysis (Homer et al. 1997)
AVHRR	AVHRR	Continuous	1000 m	NOAA (June 1990)
TM	-	-	-	Landsat TM (June 1990/August 1991)
	TM3	Continuous	30 m	TM Band 3 (Red)
	TM4	Continuous	30 m	TM Band 4 (Near-infrared)
	TM5	Continuous	30 m	TM Band 5 (Mid-infrared)

rithm for the continuous data (DMA data, precipitation, temperature, AVHRR, and TM data), and the nearest neighbor algorithm for the discrete data (geology, the classified cover-map, and district), in order to correspond with the resolution of the forest inventory data (ArcInfo GIS, ESRI Inc., Redlands, California).

Model development and selection

The 447 model-building points were intersected through each digital explanatory layer and the value at each cell extracted for use in modelling. The S-plus (StatSci Division, 1700 Westlake Ave. N., Suite 500, Seattle WA 98109) GAM function was used to generate relationships between each response variable (Table 1) and the explanatory variables (Table 2) according to the following specifications. For forest and lodgepole presence, a logit link was used to transform the mean of the response to a binomial scale. For the continuous variables, the Poisson link was used to transform the data to the scale of the response. A Poisson link was selected after evaluation of mean-variance relationships for each continuous response variable. A loess smoothing function (Venables & Ripley 1997) was chosen to summarize the relationship between the predictors and the response. The loess smoother fits a robust weighted linear function to a specified window of data. In this study, the default (0.5) window size was arbitrarily set for all smoothed functions.

Partial residuals were graphically explored for unusual patterns and outliers and the major outliers were removed from the analyses. The functional relationships between each explanatory variable and the respective response variables were then analyzed for potential parametric fits following advice of Hastie & Tibshirani (1990) and Yee & Mitchell (1991). If a potential parametric fit existed, piecewise and second- and third-order polynomial functions were fitted to the data and assessed from the relative degree of change to the residual deviance (Cressie 1991). The piecewise functions require a pre-chosen placement of ’knots’ or breakpoints within the range of the data at points where the relationships distinctively changed. The knots split the data into separate sections. A regression model is fitted to each piece of data and joined at each knot (Chambers & Hastie 1992). For this analysis, only variables with one distinctive breakpoint were fitted, with the node specified from graphical characteristics.

All explanatory variables, including all potential parametric fits, were run through a stepwise procedure to determine the best-fit model for prediction (see Chambers & Hastie 1992) using Akaike’s Information Criterion (AIC) (Akaike 1973). To examine the effects of different sources of satellite data, three stepwise procedures were performed for each forest attribute, each having the same set of explanatory variables but with a different type of satellite data. One limitation of smoothed functions obtained from GAMs is their inability to extrapolate outside the range of the data used to build the model. Therefore, values of the validation data set that were outside the range of the model-building data set were assigned the maximum/minimum value of the respective variable in the model-building data set.

Table 3. Best-fit models (bold) by satellite imagery type for predicting forest and lodgepole pine presence in the Uinta Mountains, Utah, USA. See Table 2 for variable descriptions.

Predictor variables	Forest presence			*Pinus contorta* presence		
	TM	AVHRR	TM-classified	TM	AVHRR	TM-classified
AIC	**164.7**	199.2	169.1	**198.9**	272.6	210.1
TM3	-	N/A	N/A	-	N/A	N/A
TM4	-	N/A	N/A	poly(3)	N/A	N/A
TM5	lo	N/A	N/A	-	N/A	N/A
AVHRR	N/A	-	N/A	N/A	-	N/A
GAPveg	N/A	N/A	+	N/A	N/A	+
Elev	trpw	trpw	poly(2)	poly(2)	poly(2)	poly(2)
Slp	-	-	trpw	-	-	-
Asp1	-	-	-	-	-	-
Asp2	-	-	-	-	-	-
Asp3	-	-	-	-	-	-
East	-	-	-	-	lo	lo
North	-	-	-	-	-	-
Precip	-	lo	-	-	-	-
Geol(T)	-	-	-	-	-	-
Geol(N)	+	+	-	-	-	-
Geol(R)	-	-	-	+	+	+
District	-	-	-	-	-	-

poly = polynomial of order specified in parentheses; trpw = piecewise polynomial; lo = loess smoothing function with default window span of 0.5; + = significant relationship; - = non-significant relationship.

Model validation

An independent set of data was collected in the field and compared to model predictions using error matrix analyses for the discrete, binomial responses (forest and lodgepole presence), and root mean square error (RMSE) estimations for the continuous responses (basal area, percent shrub, and snag density). RMSE was chosen as a measure that combines both bias and variance in the estimates, presented in units that have meaning to map users. A systematic grid of 3000 m intervals was applied to the Evanston District and used to select validation sites. A 3000 m interval was selected as the maximum amount of data that could be collected during one field season. The grid was randomly placed within the district boundary and field validation data collected from 96 points using standard FIA plot design and measurement procedures (Anon. 1994).

The proportion correctly classified (PCC) was calculated by dividing the sum of the diagonal values of the error matrix by the total points analyzed. A measure of randomness, the kappa statistic (KHAT) (Cohen 1960), was calculated to evaluate the effects of omission and commission errors. KHAT ranges from – ∞ to 1, with more accurate values closer to 1 and more ‘confused’ values closer to – ∞. Output from the binomial response models was a probability value scaled from 0 to 1 for each grid cell, with predictions closer to 1 indicating a greater chance of forest and lodgepole presence. *Z*-tests (corrected for multiple comparison with the Bonferroni method) were used to test for significant differences in PCC and KHAT values obtained using different satellite data as predictor variables. For the continuous response models, scatterplots were generated of field vs. predicted values to show, visually, the distribution of error, and a RMSE was calculated as:

$$RMSE = \sqrt{\sum \left(predicted - observed\right)^2 \Big/ n}. \qquad (1)$$

Predicted values within ±15% of field values were considered accurate and used to estimate PCC.

Results

Model development and selection

Binary responses

For the forest and lodgepole responses, the models including TM data had the lowest AIC value (Table 3). Both TM and TM-classified data were significant contributors to the forest and lodgepole presence models, whereas the AVHRR variable was excluded from each selected model (Table 3). Elevation and geology were selected as significant predictors in all models of forest and lodgepole presence except for the TM-classified, forest presence model, where geology was replaced with the slope parameter. Other significant variables included precipitation in the forest presence models and the UTM easting variable in the lodgepole presence models (Table 3). For the forest presence response, the TM model was similar to the AVHRR model, except that precipitation was replaced by TM Band 5 (mid-

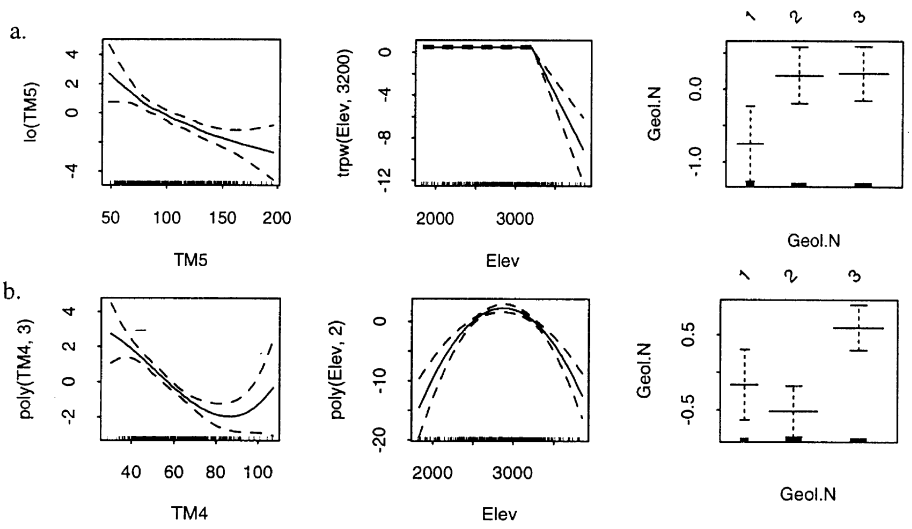

Fig 1. Explanatory variables selected from stepwise procedures as significantly contributing to the respective binomial response variables (see Tables 3 and 4 for definitions). Each plot shows the relationship of the fitted function to the response and scaled to zero. The plots include approximate 95% pointwise SE bands. At the base of eachplot is a univariate histogram (rugplot) showing the distribution of each observation. (a) Forest presence TM model. (b) Lodgepole presence TM model.

infrared) (Table 3).

The TM-classified model was similar to the lodgepole AVHRR model, but had a slightly better fit (Table 3). The primary difference between the two models was the replacement of the UTM easting variable in the AVHRR model by the TM Band 4 (near-infrared) variable in the TM model (Table 3). The probability of lodgepole cover was found to be highest at decreasing values of TM Band 4 data, elevations between 2500 and 3200 m, and on alluvial substrates (Fig. 1b).

Continuous responses

For all continuous responses except the AVHRR snag density model, all variables were selected as additively contributing to the model predictions (Table 4, Fig. 2). For the snag density model, the only variable not included was geology. As with the binomial response models, both parametric and smoothed functions were significant in each model, with models based with TM data having the lowest AIC values.

Table 4. Best-fit models (AIC and D^2) by satellite type for predicting basal area, % shrub cover, and snag density in the Uinta Mountains, Utah. Variable names are described in Table 2.

Predictor	Basal area			% Shrub cover			Snag density		
variables	TM	AVHRR	TM-classified	TM	AVHRR	TM-classified	TM	AVHRR	TM-classified
AIC	**8618.4**	11198.7	9061.9	**2983.9**	3141.7	3085.1	**4263.5**	4640.7	4606.3
D^2	43.3	29.6	**45.0**	30.7	30.1	**32.1**	**43.5**	39.9	41.1
TM3	lo	N/A	N/A	lo	N/A	N/A	lo	N/A	N/A
TM4	lo	N/A	N/A	lo	N/A	N/A	lo	N/A	N/A
TM5	lo	N/A	N/A	lo	N/A	N/A	lo	N/A	N/A
AVHRR	N/A	lo	N/A	N/A	lo	N/A	N/A	lo	N/A
GAPveg	N/A	N/A	+	N/A	N/A	+	N/A	N/A	+
Elev	trpw	trpw	trpw	lo	lo	lo	poly(3)	poly(3)	poly(3)
Slp	trpw	lo	lo	lo	lo	lo	lo	lo	lo
Asp1	lo	-	lo	poly(2)	poly(2)	poly(2)	lo	lo	lo
Asp2	-	+	-	-	-	-	-	-	-
Asp3	-	-	-	-	-	-	-	-	-
East	lo	poly(3)	lo	poly(3)	poly(3)	lo	poly(3)	poly(3)	poly(3)
North	lo	poly(3)	poly(3)	lo	poly(3)	lo	lo	lo	lo
Precip	poly(2)	poly(2)	poly(2)	poly(2)	lo	poly(2)	lo	lo	poly(2)
Geol(T)	-	-	+	-	-	-	-	-	+
Geol(N)	+	+	-	+	+	+	-	-	-
Geol(R)	-	-	-	-	-	-	+	-	-
District	+	+	+	+	+	+	-	+	+

poly = polynomial of order specified in parentheses; trpw = piecewise polynomial; lo = loess smoothing function with default window span of 0.5; + = significant relationship; - = non-significant relationship.

The relationship of elevation to basal area and snag density corresponded with the probability of forest and lodgepole presence, with high values peaking between 2500 and 3200 m, whereas shrub cover gradually declined with increasing elevations (Fig. 2). TM Bands 3 and 4 followed similar trends for each continuous response, while basal area increased and snag density slightly decreased with declining values of TM Band 5. The relationship of slope with basal area and snag density followed similar decreasing patterns, whereas the relationship of slope with shrub cover showed an initial increase up to 18% (Fig. 2). Precipitation tended to have a greater positive effect on basal area than on shrubs and snags. Basal area was greatest on alluvial substrates, shrub cover greatest on shale substrates, and snag density greatest on sedimentary rock types.

Basal area was higher at the northern and southern extremes of the mountain range, shrub cover higher at mid-latitude zones of the mountain range and snag density higher on the western edge of the range. Basal area was high in districts on the north slope, shrub cover high in Mountain View, Flaming Gorge, and Vernal districts, and snag density high in Kamas and Duchesne districts on the western end of the range.

Validation

Accuracy of the models predicting forest and lodgepole presence was high (Table 5). Differences in accuracy were not significant among the three models for either variate. RMSE values for estimates of basal area ranged from 13.9 m 2 /ha for the TM-classified model to 16.0 m^2 /ha for the AVHRR model (Fig. 3a). Sixty-three percent of the points fell within ±15% (11.5 m^2 /ha) of the true value for the TM model, 55% for the AVHRR model, and 67% for the TM-classified model. There was little difference between RMSE values for the models predicting shrub cover, with values averaging 13.8%. Seventy-five percent of the points fell within ±15% of the true cover using TM data, 77% for the AVHRR model, and 75% for the TM-classified model (Fig. 3b.). RMSE for snag density ranged from 18.1 snags for the TM model to 20.2 snags for the AVHRR model (Fig. 3c). Forty-nine percent of the points fell within ±15% of the true snag count using TM data, 41% including AVHRR data, and 54% with TM-classified data.

Table 5. Percent correctly classified (PCC) and estimates of Kappa (KHAT) for TM, AVHRR and TM-classified models predicting forest and lodgepole pine presence in the Uinta Mountains, Utah. Bold-faced values indicate highest accuracy.

Forest attribute	Satellite type	PCC	KHAT
Forest	TM	**86.5**	**0.58**
presence	AVHRR	82.3	0.43
	TM-classified	85.4	0.54
Pinus contorta	TM	71.9	0.38
presence	AVHRR	71.9	0.37
	TM-classified	**80.2**	**0.56**

Discussion

Generalized Additive Models

Clearly, vegetation communities do not exhibit 'normal' (Gaussian) distribution patterns throughout the environment (Austin et al. 1990, 1994); therefore, predictability is dependent on the flexibility and capability of the analytical procedures used to model vegetation distribution. GAMs, in contrast to some analytical procedures (e.g., ordination and linear regression models), do not make a priori assumptions about underlying relationships, thus allowing the data to drive the fit of the model instead of the model driving the data. The graphical nature of GAMs also allows for the opportunity to visualize the additive contribution of each variable to the respective response using smoothed functions (Figs. 2, 3). Smoothed functions are capable of fitting unusual variance patterns such as skewness and bimodality that are often overlooked with standard linear models (Austin & Noy-Meir 1971). A limitation of GAMs we encountered in this study was the uncertainty associated with extrapolation of the smoothed functions, particularly at the tails of the distribution. As suggested by Hastie & Tibshirani (1990) and Yee & Mitchell (1991), we fitted parametric functions to the model whenever 'statistically allowable', thus constraining the behavior of the functions in the extreme ranges of the data. Often this involved a subjective interpretation based on visual inspection of the data.

We found GAMs to be powerful exploratory tools for detecting simple linear relationships as well as complex patterns in forest attribute distribution, and tools flexible enough for integrating both parametric and nonparametric functions in the models. For example, most of our models included at least one smoothed function as a predictor variable, indicating a better model fit was achieved using a nonlinear distribution. This supports findings of other studies (Austin & Cunningham 1981; Austin 1987; Margules & Stein 1989; Leathwick & Mitchell 1992), where relationships of environmental variables to plant species' responses were not always best described by Gaussian distributions.

Elevation was a significant predictor in all models. This is not surprising in a mountainous environment like Utah, where elevation, a surrogate for moisture and temperature gradients (Barbour et al. 1987), is a driving mechanism for vegetation distributions. The limitation of using an indirect variable, such as elevation, as a predictor variable is that the vegetation response is limited to the characteristics of the species' local environment (Austin et al. 1984, Austin & Smith 1989). Model effectiveness may therefore be limited when applied to environments outside the range where the model was developed.

The forest presence models indicated the additive importance of geologic features (nutrients) and moisture variables, such as total annual precipitation and the spectral signatures of moisture (TM Band 5), along with elevation. This is not surprising given that the essential environmental gradients influencing vegetation production are moisture, temperature, and nutrients (Barbour et al. 1987). The difference between the lodgepole and forest presence models was the added significance of geographic location (UTM Easting coordinates) and Band 4 (near-infrared) of the TM data in predicting lodgepole presence. TM Band 4, which discriminates green biomass, was a better predictor for lodgepole than the moisture-related TM spectral Band 5 (mid-infrared). This suggests the importance of spectral data for discriminating highly disturbed or successional-stage forest types, such as lodgepole pine.

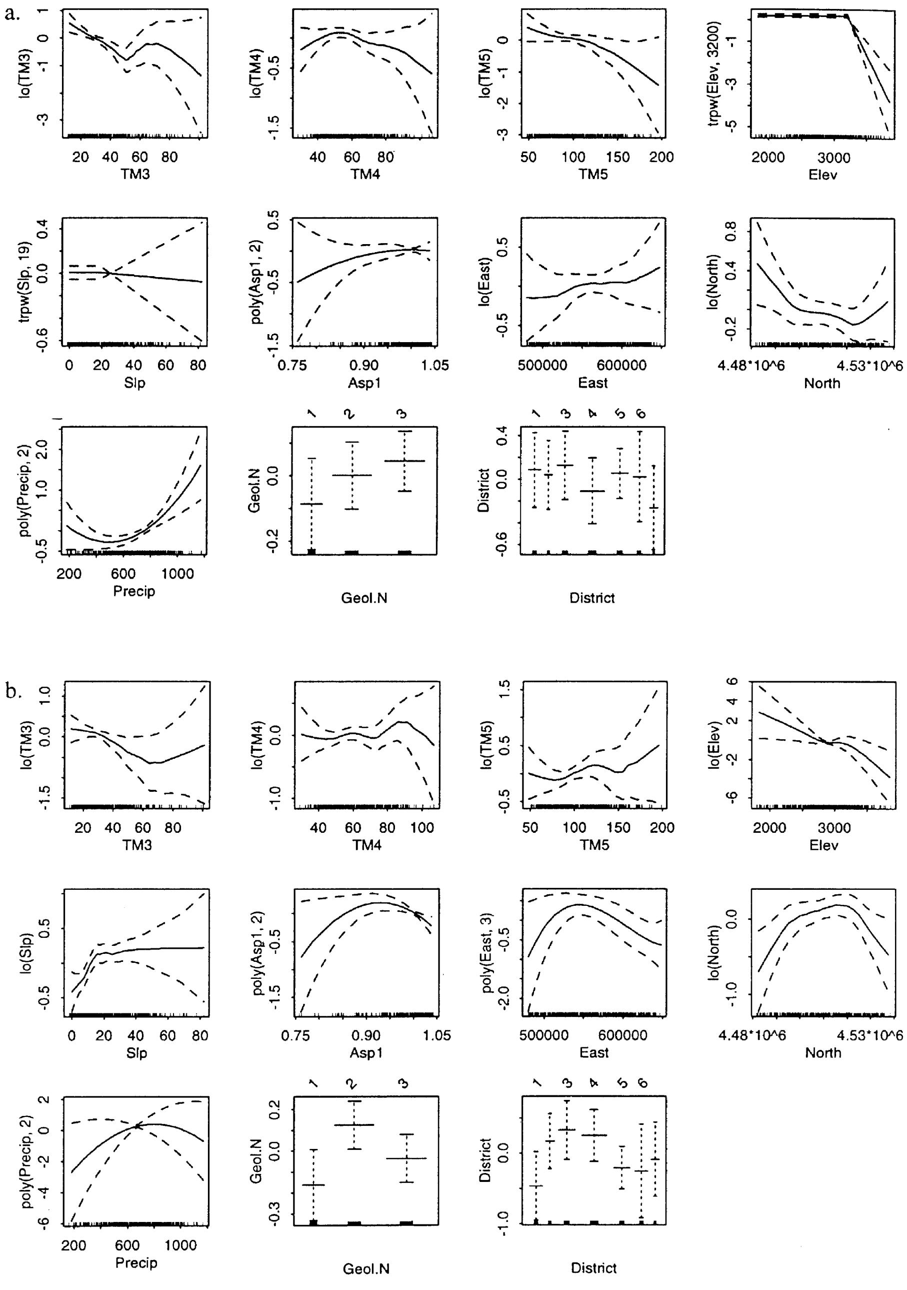

Fig. 2. Explanatory variables selected from stepwise procedures as significantly contributing to the respective binomial response variables (see Tables 3 and 4 for definitions). Each plot shows the relationship of the fitted function to the response and scaled to zero. The plots include approximate 95% pointwise SE bands. At the base of eachplot is a univariate histogram (rugplot) showing the distribution of each observation. (a) Basal area. (b) % shrub cover. (c) Snag density. see next page.

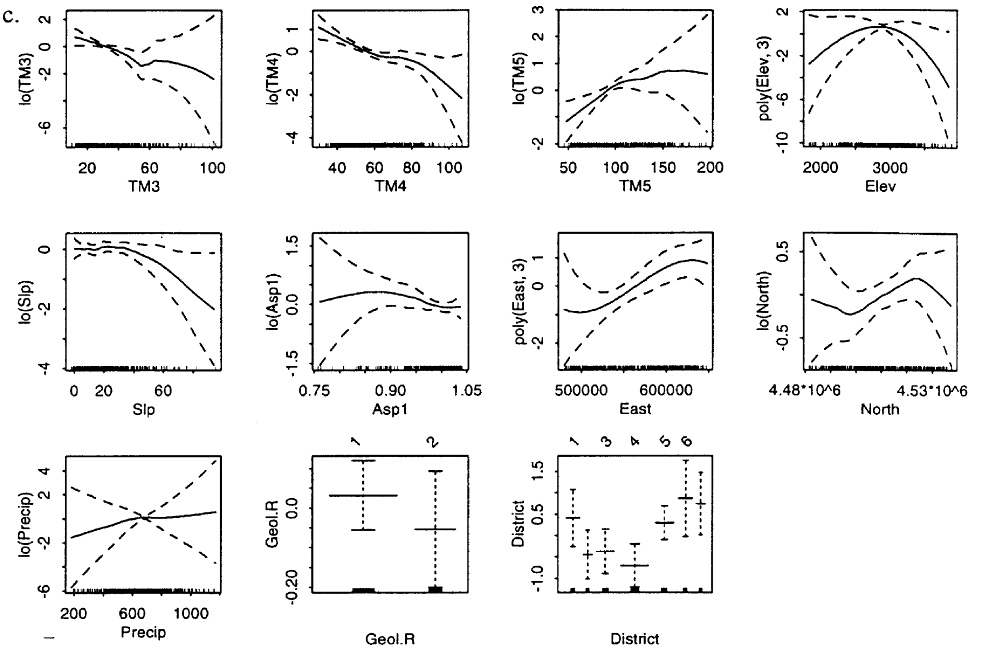

Fig. 2. *(Cont.)* Explanatory variables selected from stepwise procedures as significantly contributing to the respective binomial response variables (see Tables 3 and 4 for definitions). Each plot shows the relationship of the fitted function to the response and scaled to zero. The plots include approximate 95% pointwise SE bands. At the base of eachplot is a univariate histogram (rugplot) showing the distribution of each observation. (a) Basal area. (b) % shrub cover. (c) Snag density.

The distribution of basal area, shrub cover, and snag density within the Uinta Mountain range appeared to be related to all environmental variables specified in the initial model. Questions remain on the magnitude of forest attribute response to each environmental gradient and whether variables not represented in this study are affecting vegetation structure. Other considerations include the impact of the human population on forest diversity. Human intervention has introduced fragmentation of vegetation communities from roads and clearcuts, and extensive habitat and diversity loss from human development, timber management, wildfire suppression, and livestock grazing. These disturbances strongly effect forest composition and can weaken the relationship between predicted and actual forest attributes.

Validation

Assessing model accuracy was not without questions. For validating discrete data sets, PCC provides a measure of overall accuracy, but does not provide information about omission and commission errors included in the predictions. This study included coinciding Kappa (KHAT) values, which provide a measure of improvement of the model over random predictions, incorporating omission and commission errors (Cohen 1960). In general, the accuracy of forest and lodgepole presence models was high, with PCC ranging from 82.3% to 86.5% and 71.9% to 80.2%, respectively. These values are well within the range of accuracies estimated for discrete cover-types (Edwards et al. 1998).

Error matrix calculations work well for discrete data types, but are not appropriate for analyzing continuous data. RMSE provided an estimate of model variance, averaging 14.7 m^2 /ha for basal area, 13.8% for shrub cover, and 19.0 for number of snags. The scatterplots of field reference vs. prediction displayed the distribution of error in the data. In general, the models tended to underpredict at high values of basal area, shrub cover, and snag density and overpredict at locations sampled as having no forest cover (Fig. 3). This bias between observed and predicted values may be caused by the influence of ‘naughty noughts’, or zero values which, when large numbers of zero values are included in the model building data set, tend to distort the shape of the response function (Austin & Meyers 1996). Overdispersed data with additional zeros (from whatever cause) appear common in ecological data sets; further research on this topic is needed from both an ecological and a statistical perspective (See Austin & Cunningham 1981; Austin et al. 1994; Austin & Meyers 1996 for examples).

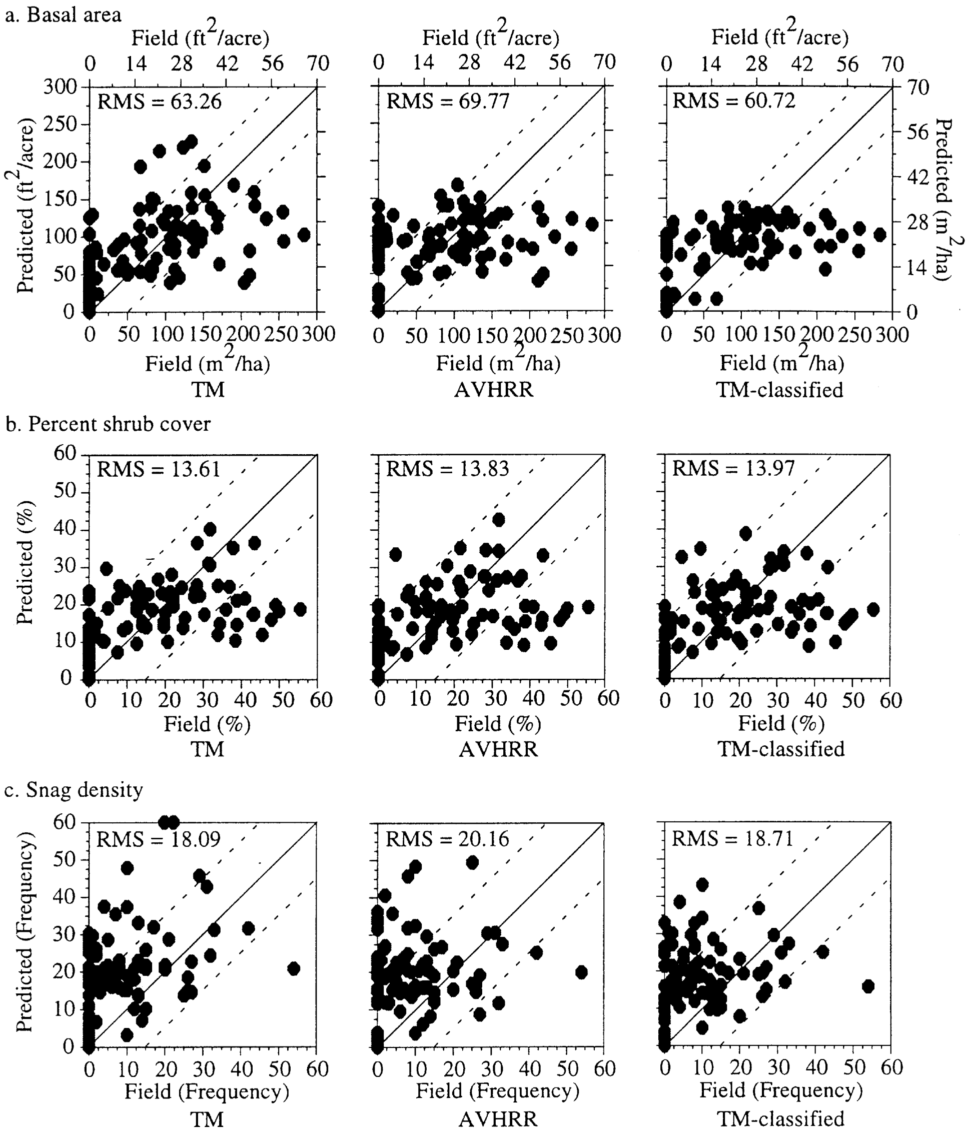

Fig. 3. Scatterplots of field reference data vs. model predictions, including RMS values. The solid lines represent perfect correlation between the predicted and reference values, and the dotted lines show user-defined acceptable deviations fromperfect correlation. (a) Basal area with reference lines at +/– 11.5 m^2/ha. (b) % shrub cover with reference lines at +/– 15%. (c) snag density with reference lines at +/– 15 snags.

Satellite data

A satellite data component was selected as significant in all models except the forest and lodgepole presence models including AVHRR. This supports findings that satellite data used in conjunction with environmental digital data enhances model predictions (Strahler et al. 1979; Davis et al. 1991). The AVHRR component did not contribute to the model-building process as much as the TM-classified or TM data (Table 3, Table 4), and had lower accuracy when compared with our validation data set (Table 5, Fig. 3). Spectral values influenced by shadows or extreme moisture differences may actually detract from useful information for prediction. In a classified map, these values are discriminated by ecological characteristics and nearby pixels and therefore enhance information extraction from the raw spectral data. This may be a reason why the TM models in most cases were less accurate than the TM-classified models. Also, only three TM spectral bands (3, 4, and 5) were included in the TM models for this study, whereas the TM-classified cover map included all six bands (1, 2, 3, 4, 5, and 7) for classification procedures. Questions remain on the effects of using different bands or combinations of bands in the model. For example, Franklin (1986) found significant relationships between visible reflectance bands (Bands 1, 2, and 3) and stand basal area and leaf biomass for coniferous vegetation while Ahern (1992) found significant relationships between bands 7/4 and spruce-fir volume. Unfortunately, GAMs are not effective at high dimensions, and modelling approaches examining many variables simultaneously, such as would be needed to analyze the interaction among many different spectral bands, should be explored with caution.

Acknowledgements. Funding for this research was provided by the USDA Forest Service, Rocky Mountain Research Station, Ogden, Utah, in cooperation with the USGS Biological Resources Division, Utah Cooperative Fish and Wildlife Research Unit, Utah State University, and the USGS Biological Resources Division Gap Analysis Program. We would like to thank D. Roberts, D. Chojnacky, N. Zimmermann, J. Kershner, and two anonymous reviewers for their constructive comments on this and other earlier drafts. Additional thanks are given to our field assistants, 'Super Sue' Russell, T. B. Murphy, C. Isom, D. Toone, J. Lawler, J. DeNormandie, and the folks at Forest Survey, in particular D. Van Hooser, for their support and enthusiasm.

References

Anon. 1994. *Utah forest Survey field procedures, 1994-1995.* Unpublished field guide on file at: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Forestry Sciences Laboratory, Interior West Resource Inventory, Monitoring, and Evaluation Program, Ogden, UT.

Ahern, F.J. 1992. Satellite data applied to forest management. *Rem. Sens. Can.* 20: 4-5.

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In: Petran, B.N. & Csàki, F. (eds.) *International Symposium on Information Theory. 2nd ed.*, pp. 267-281. Akadémiai Kiadi, Budapest.

Austin, M.P. 1985. Continuum concept, ordination methods and niche theory. *Annu. Rev. Ecol. Syst.* 16: 39-61.

Austin, M.P. & Noy-Meir, I. 1971. The problem of non-linearity in ordination: experiments with two gradient models. *J. Ecol.* 59: 762-773.

Austin, M.P. & Cunningham, R.B. 1981. Observational analysis of environmental gradients. *Proc. Ecol. Soc. Aus.* 11: 109-119.

Austin, M.P. & Meyers, J.A. 1996. Current approaches to modelling the environmental niche of eucalypts: implication for management of forest biodiversity. *For. Ecol. Manage.* 85: 95-106.

Austin, M.P. & Smith, T.M. 1989. A new model for the continuum concept. *Vegetatio* 83: 35-47.

Austin, M. P., Cunningham, R.B. & Fleming, P.M. 1984. New approaches to direct gradient analysis using environmental scalars and statistical curve-fitting procedures. *Vegetatio* 55: 11-27.

Austin, M.P., Nicholls, A.O. & Margules, C.R. 1990. Measurement of the realized qualitative niche: environmental niches of five *Eucalyptus* species. *Ecol. Monogr*. 60: 161-177.

Austin, M.P., Nicholls, A.O., Doherty, M.D. & Meyers, J.A. 1994. Determining species response functions to an environmental gradient by means of a beta function. *J. Veg. Sci.* 5: 215-228.

Barbour, M.G., Burk, J.H. & Pitts, W.D. 1987. *Terrestrial plant ecology*. The Benjamin/Cummings Publishing Co. CA.

Chambers, J.M & Hastie, T.J. 1992. *Statistical models in S.* Wadsworth and Brooks/Cole, Pacific Grove, CA.

Cohen, J. 1960. A coefficient of agreement of nominal scales. *Educ. Psychol. Meas.* 20: 37-46.

Cressie, N.A.C. 1991. *Statistics for spatial data.* J. Wiley, New York, NY.

Cronquist, A., Holmgren, A.H., Holmgren, N.H. & Reveal, J.L. 1972. *Intermountain flora (Vol. 1): Vascular plants of the Intermountain West, U.S.A.* Hafner, New York, NY.

Daly, C., Nielson, R.P. & Phillips, D.L. 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteorol.* 33: 140-158.

Davis, F.W. & Goetz, S. 1990. Modeling vegetation pattern using digital terrain data. *Landscape Ecol*. 4: 69-80.

Frank, T. 1988. Mapping dominant vegetation communities in the Colorado Rocky Mountain front range with Landsat Thematic Mapper and digital terrain data. *Photogramm. Eng. Remote Sens.* 54: 1727-1734.

Franklin, J. 1986. Thematic mapper analysis of coniferous forest structure and composition. *Int. J. Remote Sens.* 7: 1287-1301.

Frescino, T.S. 1998. *Development and validation of forest habitat models in the Uinta Mountains, Utah.* M.Sc. Thesis, Utah State University, Logan, UT.

Hastie, T.J. & Tibshirani, R.J. 1990. *Generalized Additive Models*. Chapman and Hall, London.

Hintze, L.F. 1980. *Geologic map index of Utah*. Utah Geological and Mineralogical Survey, Salt Lake City, UT.

Homer, C.G., Ramsey, R.D., Edwards, T.C. Jr. & Falconer, A. 1997. Landscape cover-type modelling using a multi-scene thematic mapper mosaic. *Photogramm. Eng. Remote Sens.* 63: 59-67.

Horler, D.N.H. & Ahern, F.J. 1986. Forestry information content of Thematic Mapper data. *Int. J. Rem. Sens.* 7: 405-428.

Leathwick, J.R. & Mitchell, N.D. 1992. Forest pattern, climate and vulcanism in central North Island, New Zealand. *J. Veg. Sci.* 3: 603-614.

Loveland, T.R., Merchant, J.W., Ohlen, D.O. & Brown, J.F. 1991. Development of a land-cover characteristics database for the conterminous U.S. *Photogramm. Eng. Remote Sens.* 57: 1453-1463.

Mauk, R.L. & Henderson, J.A. 1984. *Coniferous forest habitat types of northern Utah.* USDA For. Serv. Gen. Tech. Rep. INT-170, Ogden, UT.

Margules, C.R. & Stein, J.L. 1989. Patterns in the distributions of species and the selection of nature reserves: an example from Eucalyptus forests in south-eastern New South Wales, Australia. *Biol. Conserv.* 50: 219-238.

Moisen, G.G. & Edwards, T.C., Jr. 1999. Use of generalized linear models and digital data in a forest inventory of northern Utah. *J. Agric. Biol. Environ. Statist.* 4:164-182.

Mueller-Dombois, D. & Ellenberg, H. 1974. *Aims and methods of vegetation ecology*. Wiley, New York, NY.

Roberts, D.W. & Cooper, S.V. 1989. Concepts and techniques of vegetation mapping. In: Ferguson, D., Morgan, P. & Johnson, F.D. (eds.) *Land classifications based on vegetation: applications for resource management*, pp. 90-96. USDA For. Serv. Gen. Tech. Rep. INT-257, Ogden, UT.

Stenback, J.M. & Congalton, R.G. 1990. Using Thematic Mapper imagery to examine forest understory. *Photogramm. Eng. Remote Sens.* 56: 1285-1290.

Strahler, A.H., Logan, T.L. & Woodcock, C.E. 1979. Forest classification and inventory system using Landsat, digital terrain, and ground sample data. In: *13th International Symposium on Remote Sensing of Environment*, pp. 1541-1557. Environmental Research Institute of Michigan, Ann Arbor, MI.

Swift, L.W. Jr. 1976. Algorithm for solar radiation on mountain slopes. *Water Resourc. Res.* 12: 108-112.

Venables, W.N. & Ripley, B.D. 1997. *Modern applied statistics with S-plus*. Springer-Verlag, New York, NY.

Yee, T.W. & Mitchell, N.D. 1991. Generalized additive models in plant ecology. *J. Veg. Sci.* 2: 587-602.

Received 19 October 1999;
Revision received 27 March 2000;
Final revision received 20 June 2000;
Accepted 20 June 2000.
Coordinating Editor: J.B. Wilson.